引用格式：郭华东，陈和生，闫冬梅，等. 加强开放数据基础设施建设，推动开放科学发展. 中国科学院院刊, 2023, 38(6): 806-817

Guo H D, Chen H S, Yan D M, et al. Strengthening open data infrastructure and promoting open science. Bulletin of Chinese Academy of Sciences, 2023, 38(6): 806-817

加强开放数据基础设施建设，推动开放科学发展

郭华东[1,2*]　陈和生[3]　闫冬梅[1,2]　黎建辉[1,4]　陈　刚[3]　何国金[1,2]　马俊才[5]　梁　栋[1,2]　孔玲贻[1,2]

1 可持续发展大数据国际研究中心　北京　100094
2 中国科学院空天信息创新研究院　北京　100094
3 中国科学院高能物理研究所　北京　100049
4 中国科学院计算机网络信息中心　北京　100083
5 中国科学院微生物研究所　北京　100101

摘要　为促进经济发展、社会进步与科技创新，必须大力推动开放科学，开放数据成为开放科学的重要组成。开放数据基础设施由物理实体和虚拟系统组成，是支撑开放数据和满足不同领域数据应用需求的基础，其发展是信息技术演进的客观过程。文章分析了开放数据基础设施的要素，阐明了开放数据基础设施对促进开放科学实施的意义和积极作用，分析了我国开放数据基础设施的现状，针对我国开放数据基础设施存在的不足与挑战提出了发展建议。

关键词　开放数据，开放数据基础设施，开放科学，地球大数据，科学数据中心

当前，开放科学进入全球共识阶段，科研人员基于活动论、方法论、方式论、过程论、文化论和实践论等从不同视角对开放科学给出了不同的定义[1]。开放科学将科学作为一种共同利益加以推广，包括分享数据、方法、结果和由此产生的知识等。作为一种全新的知识体系，开放科学强调整个科学过程的透明，鼓励开放获取与合作。在更深层次的理解上，开放科学是科学研究范式的转变[2,3]，将深刻改变人类科学研究、科学发现的方式，对于加强科技合作，共同探索解决全球性挑战具有重要意义[4]。

* 通信作者
资助项目：中国科学院学部咨询评议项目（2022-X01-B-008、2022KJC-Y-0315）
修改稿收到日期：2023 年 6 月 7 日

为推动全球开放科学的发展，2021 年，联合国教科文组织（UNESCO）大会第 41 届会议审议通过《开放科学建议书》（*Recommendation on Open Science*）[5,6]，意在让开放科学更透明、更可及，进而使其更加公平和包容。这不仅为开放科学提供了国际发展方向，还为个人、机构、国家、地区和国际等不同层面提供了促进交流、建立信任的渠道。

近年来，随着大科学装置的建设、重大科学实验的实施，以及科学传感器和传感网络的广泛应用，产生了多源、异构、海量的科学数据。数据已不仅是研究过程中事实或观察的结果，更成为研究的实体工具。数据密集型科学发现成为大数据时代的科学研究新范式[7]。

开放科学数据是开放科学的核心要素之一[2]。通常由政府和机构组织制定开放数据政策，定义共享数据类型、共享对象、共享条件[8]。实现科学数据的公开使用、重复使用、长期保存和更新发布，将会大力促进开放科学的发展，提升科技创新活动的透明度、可重复性、协作性，并可最大化科学对社会发展的价值[9]。

1 开放科学数据

科学数据，作为最基本、最活跃的一类科技资源，既是科技创新活动的重要产出，也是新一轮创新活动和经济社会发展的重要基础，具有巨大的应用和开发价值，在全球诸多国家机构得到高度重视。

开放科学数据是指可公开获取，并可通过下载、复制、分析、再加工等方式，用于系统构建和任何其他应用目的的科学数据[10-12]。

1.1 国际开放科学数据的发展与现状

在美国，开放科学数据已成为美国“信息自由、开放政府”的重要组成；德国科学基金会（DFG）2010 年发布的《研究数据操作指南》（*Guidelines on the Handling of Research Data*），明确“科学数据的共享与再利用对科学研究甚至人类社会都具有巨大意义”[13]。国际科学理事会数据委员会（CODATA）2019 年发布的《科研数据北京宣言》（*The Beijing Declaration on Research Data*）[14]指出，公共经费资助产出的科学数据应尽可能在全球范围内共享重复使用。以 Dimensions 平台统计数据为例，2012—2021 年全球开放科学数据集合计 9 918 741 条，其中中国拥有 253 441 条，继美国之后排名全球第 2 位；在研究领域上，数据量排名前 3 位的学科分别为信息与计算机科学、信息系统、地球科学（图 1）。

为推进开放数据更加规范化，2016 年 FAIR 原则[15,16]——可查找（findable）、可获取（accessible）、可交互（interoperable）、可重复使用（reusable）应

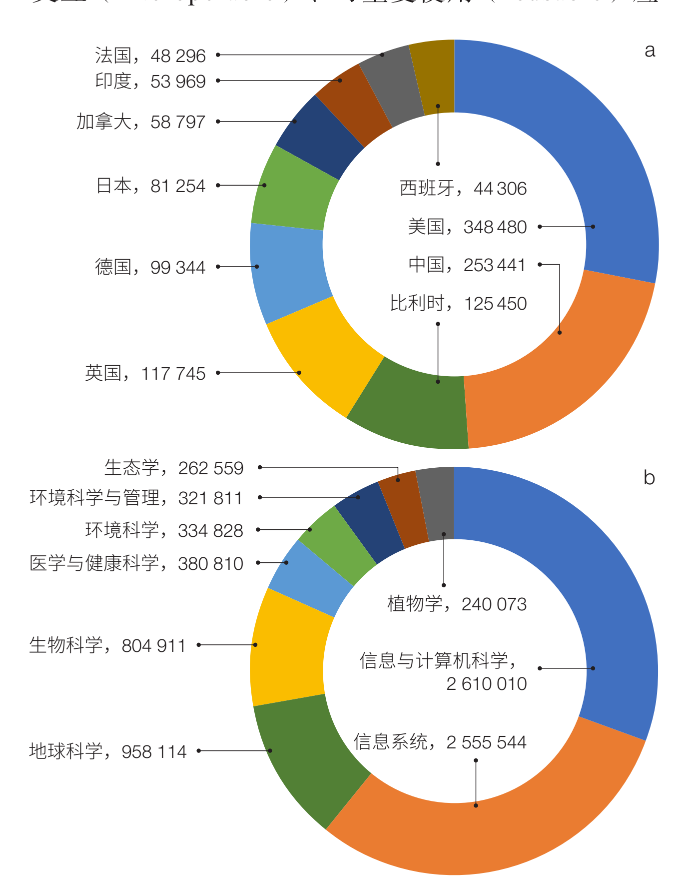


图 1 2012—2021 年全球开放科学数据集分布信息

Figure 1 Distribution of open scientific dataset globally from 2012 to 2021

（a）不同国家数据集分布信息；（b）不同学科数据集分布信息

(a) Distribution of different countries; (b) Distribution of different disciplines

数据来源：Dimensions 平台

Data resource: Dimensions platform

运而生。作为FAIR原则的补充，CARE原则——集体收益（collective benefit）、质量保证（authority to control）、责任（responsibility）、伦理（ethics）则倡导以目标为导向，发挥数据的创新作用。FAIR原则强调技术进步，而CARE原则更侧重政策变革，两者相辅相成[17]。

全球开放数据在众多研究领域蓬勃发展。作为开放科学的引领者，天文学领域越来越多的数据资源实现了即时开放，如美国大型综合巡天望远镜（LSST）。在空间科学领域的探索中，美国和加拿大最为积极，欧洲以试点为主，俄罗斯和日本则强调国际合作；在高能物理领域，依托大科学装置的数据积累与协作管理经验为全球大规模数据合作共享提供了典范。

1.2 我国开放科学数据发展与现状

作为国家科技创新发展和经济社会发展的重要基础性战略资源，开放科学数据已在我国全社会达成高度共识。

1.2.1 历史上形成了丰富的、覆盖各个领域的科学数据资源

据不完全统计，到20世纪末，我国已建成5 000—6 000个规模不等、质量参差的科学数据库，涉及科学技术的各个领域。在科学数据采集和积累方面，初步形成了以部门为主体、科研院所和高等院校互补的格局，并逐步建立了专门的数据管理机构和国际数据合作与交换渠道[18]。

据《国家科学数据资源发展报告（2018）》统计，截至2017年底，我国有效管理与保存的科学数据资源总量共计约83.72 PB。其中，生命科学与医学领域、地球与环境科学领域、物理与化学领域、对地观测领域、天文与空间科学等5个领域数据积累分别为26.81 PB、24.48 PB、16.64 PB、9.73 PB和5.27 PB。

1.2.2 科学数据开放共享被逐步纳入我国政策法规制度体系

我国一直重视科学数据的汇交管理与开放共享。先后从国家、部门多层面出台相关政策，并从行动上予以落实。

2006年，国务院发布《国家中长期科学和技术发展规划纲要（2006—2020年）》，明确提出要建设数字科技平台，促进科学数据共享。我国逐步开始形成由以政府、行业机构和领域数据中心为主体的数据政策体系[19]。2015年，国务院印发《促进大数据发展行动纲要》，提出发展科学大数据的具体行动。2018年，国务院办公厅印发《科学数据管理办法》，进一步明确了科学数据管理与共享的职责，并对科学数据的定义、管理、责任主体、使用方法、使用范围和保密安全进行了清晰的界定，为科学数据的开放和共享奠定了基础[20]。

1.2.3 积极参与并推动国际科学数据合作共享

当前国际科学理事会（ISC）有两大数据组织——国际数据委员会（CODATA）和世界数据系统（WDS，前身为世界数据中心WDC），主要致力于将科学技术各领域从事数据工作的科学家组织起来，利用国际网络构建全球尺度的科学数据交换体系[21,22]。1988年我国加入WDS，并于当年建立了天文、空间科学、海洋、气象、地质、地震、地球物理、冰川冻土、可再生资源与环境9个数据中心。目前，天文数据中心、空间科学数据中心和海洋数据中心均为WDS的常规成员机构。1984年我国加入CODATA。2011年，我国学者在CODATA提出“手拉手合作伙伴计划”，旨在促进国际项目之间的合作，帮助重复利用已有数据资源，减少重复劳动，提高科研成果产出，加速科研成果转化，消减数字鸿沟[21,23]。

2018年立项的中国科学院战略性先导科技专项（A类）“地球大数据科学工程”（CASEarth）遵循从开放数据到开放科学的发展潮流，将人工智能赋能地球大数据，打造集数据、计算、服务于一体的数据共享新模式[24]。同时，CASEarth促进地球科学数据的整合，实现多学科数据关联分析和信息融合，驱动重大科学发

现与决策支持，应对全球可持续发展等重大挑战[25]。

在 CASEarth 基础上，可持续发展大数据国际研究中心于 2021 年 9 月 6 日正式成立，旨在为解决中国乃至全球重大可持续发展问题提供基础理论、技术方法、决策支持和智库服务支撑[26,27]。这是我国积极推动大数据服务于《联合国 2030 年可持续发展议程》的重要举措和创新方向[28]。

1.2.4 积极创办数据期刊和创建数据仓储，推动数据开放共享

随着开放数据理念不断深入，我国也开始了科学数据期刊的建设。中国科学院计算机网络信息中心于 2016 年创办的《中国科学数据》是我国目前首批面向多学科领域的数据期刊之一[29]。生物学、遥感科学与技术和图书情报学等领域享有较高声誉的学术期刊也陆续设置了数据论文专栏[30]。由国际数字地球学会和 CASEarth 共同主办的 *Big Earth Data* 于 2017 年 12 月创刊，是全球首本聚焦大数据的地学刊物，旨在为从事地球大数据的采集、管理、处理、分析和可视化研究的学者搭建一流的国际学术交流平台[31]。

我国还在不同领域创建了数据仓储平台，来促进各领域数据资源的开放共享。国内多个数据平台还被国际认证为数据仓储中心，从而更好地保障了我国科学家的数据主权[32]。

2 我国开放数据基础设施的现状与问题

开放数据基础设施是支撑开放数据和满足不同科学领域研究的共享设施，其发展是信息技术面向数字化、智能化演进的必然结果。数据中心是开放数据基础设施的具体体现，以数据为核心，通过深度整合计算、存储、网络和软件资源，实现开放数据的价值最大化。

开放数据基础设施的要素包括数据、物理、技术、制度 4 部分。其中，数据要素是开放数据基础设施的第一要素，包括数据集、数据标识和数据注册等；物理要素是指面向数据标准、存储、管理、共享、分析、使用所需的软硬件基础设施；技术要素是面指向开放数据基础设施的核心技术研发，实现协作式和多学科数据分析的开放计算和数据处理能力；制度要素为开放数据和开放数据基础设施的使用和管理提供政策指导。

2.1 开放数据基础设施对促进开放科学实施的意义和积极作用

开放数据基础设施是科研范式变革的重要助推剂，是重要科技突破的“新引擎”[33,34]。随着互联网、大数据与人工智能的发展，科学发现的路径进入了新阶段[35]。例如，中国科学院高能物理研究所与国家高能物理科学数据中心联合研建的大规模分布式数据服务基础设施平台，通过超高速网络将国际上重要的高能物理科学数据及计算资源进行整合，为粒子物理、天体物理、中子科学、光子科学等领域的科学发现提供数据服务。高海拔宇宙线观测站项目（LHAASO）利用该平台部署了 Coryda 数据处理系统，全面收集和处理超高能伽马射线等宇宙线数据。以 2022 年为例，LHAASO 全年共采集了 11 PB 的数据，包含 10 万亿个宇宙线事例；数据和计算资源向全球开放，全年的数据访问和处理量达到 448 PB，在宇宙线前沿研究方面获得了“PeV 超高能光子”“超高质量暗物质寿命”等多项重大科技成果。

2.2 我国开放数据基础设施的基础与优势

开放数据基础设施的建设需求因所服务对象不同及应用差异而呈现多样化[36]。开放数据基础设施主要可分为两类：① 大科学装置，根据共性需求提供服务的国家基础设施，是海量数据的生产单元；② 科学数据中心，属于集中式数据基础设施，可支持跨学科领域的综合研究。中国科学院在我国开放数据基础设施建设和体系化运行中发挥了重要作用。

2.2.1 大科学装置

我国大科学装置最早可追溯到为“两弹一星”

研制任务建设的大型科研装置。20 世纪 80 年代末中国科学院率先建设北京正负电子对撞机；我国“九五”“十五”期间建设了11项大科学装置；“十一五”之后进入快速发展时期，“十二五”期间形成了建成22项、在建16项的布局；“十三五”“十四五”期间则逐步形成以综合性国家科学中心为依托的大科学装置建设规划[37]。目前，我国在建和运行的大科学装置总量约50余个，部分装置综合水平进入全球“第一方阵”[38]。

大科学装置分为三大类：① 专用研究设施，主要为特定学科领域的重大科学技术目标建设；② 公共实验设施，主要支撑多学科领域的基础和应用研究服务；③ 公益科技设施，主要为国家经济建设、国家安全和社会发展提供基础数据[39]。前两类装置一般通过实验与观测产生大量具有极高科学价值的数据，提供给专业领域和多学科用于科学研究；第三类装置通过科学考察、综合检测等手段获取科学数据及资源，为科学研究及国家发展提供保障。

大科学装置是科学数据最重要的生产源。我国与世界发达国家都高度重视大科学装置发展，但在对大科学装置的地位和作用规定的内涵的定位上有很大差别，主要体现在我国更加突出“目标牵引、问题导向”。在国家有关部门的统一部署下，我国大科学装置布局逐步完善、运行更加高效、产出更加丰硕，这对促进我国科学技术事业发展起到了巨大的支撑作用，为解决国家发展中遇到的关键瓶颈问题作出了突出贡献。

2.2.2 科学数据中心

（1）国际科学数据中心。国际科学数据中心是指面向国家和全球发展部署，服务解决重大科学问题，推动技术创新，促进可持续发展的基础设施。例如，可持续发展大数据国际研究中心（CBAS）是典型的国际科学数据中心[40]，其研发的大数据平台系统（SDGs 大数据平台）整合了基础地理、遥感、地面监测、社会统计等多种数据，贯通“大数据存储—管理—计算分析—可视化”流程；研发可持续发展目标（SDGs）数据产品生产系统，实现TB量级数据交互式在线分析，以及各类指标在线计算和可视化展示；研发SDGs专用存储库等核心功能，支持全球SDGs数据资源的持续汇聚与开放共享；建设支撑地球大数据管理、处理与分析的专有环境，具备每秒1 000万亿次的双精度浮点超级计算能力，50 PB数据存储能力，10000 CPU核心云计算能力。目前，平台已汇聚数据量达16 PB，可面向公众、科研人员、决策者三类典型场景，提供“一站式”数据计算、分析、展示、共享服务[24]，已通过CODATA的评估，科学数据服务于174个国家和地区。

（2）国家科学数据中心。2019年，为进一步完善科技资源共享服务体系，推动科技资源向社会开放共享，科学技术部、财政部联合认定了20个国家科学数据中心（表1），涵盖了高能物理、空间科学与天文、生物基因、环境与生态、地质与地震、农林、气象等领域，负责我国相关领域科学数据的汇交与共享、科学计算、数据技术研究等工作。至2021年底，国家科学数据中心汇集的数据超过100 PB，每年被访问的数据达数百PB，提供的科学计算服务超过1亿CPU小时，为科学发现、技术创新和国民经济提供了重要支撑。

（3）中国科学院科学数据中心体系。为落实《科学数据管理办法》，2019年2月中国科学院出台《中国科学院科学数据管理与开放共享办法》，并启动建设了以“总中心-学科中心-所级中心”三类科学数据中心为核心，安全体系、运行体系和评价体系共同保障与驱动的一体化科学数据中心网络[41,42]。初步建成中国科学院科学数据中心体系（表2），在支持我国科技创新方面取得了积极成效，在支撑国家重大战略、重大工程建设中发挥了积极作用。

2.3 我国开放数据基础设施发展面临的挑战

我国早期的开放数据基础设施可靠性较差，且相

表1　20个国家科学数据中心

Table 1　List of 20 National Science Data Centers

序号	国家平台名称	依托单位	主管部门
1	国家高能物理科学数据中心	中国科学院高能物理研究所	中国科学院
2	国家基因组科学数据中心	中国科学院北京基因组研究所	中国科学院
3	国家微生物科学数据中心	中国科学院微生物研究所	中国科学院
4	国家空间科学数据中心	中国科学院国家空间科学中心	中国科学院
5	国家天文科学数据中心	中国科学院国家天文台	中国科学院
6	国家对地观测科学数据中心	中国科学院空天信息创新研究院	中国科学院
7	国家极地科学数据中心	中国极地研究中心	自然资源部
8	国家青藏高原科学数据中心	中国科学院青藏高原研究所	中国科学院
9	国家生态科学数据中心	中国科学院地理科学与资源研究所	中国科学院
10	国家材料腐蚀与防护科学数据中心	北京科技大学	教育部
11	国家冰川冻土沙漠科学数据中心	中国科学院寒区旱区环境与工程研究所	中国科学院
12	国家计量科学数据中心	中国计量科学研究院	国家市场监督管理总局
13	国家地球系统科学数据中心	中国科学院地理科学与资源研究所	中国科学院
14	国家人口健康科学数据中心	中国医学科学院	中华人民共和国国家卫生健康委员会
15	国家基础学科公共科学数据中心	中国科学院计算机网络信息中心	中国科学院
16	国家农业科学数据中心	中国农业科学院农业信息研究所	农业农村部
17	国家林业和草原科学数据中心	中国林业科学研究院资源信息研究所	国家林业和草原局
18	国家气象科学数据中心	国家气象信息中心	中国气象局
19	国家地震科学数据中心	中国地震台网中心	中国地震局
20	国家海洋科学数据中心	国家海洋信息中心	自然资源部

表2　中国科学院科学数据中心体系（总中心-学科中心）

Table 2　Data Science Centers of Chinese Academy of Sciences

序号	中国科学院平台名称	依托单位	类别
1	中国科学院科学数据总中心	中国科学院计算机网络信息中心	总中心
2	11家国家科学数据中心	中国科学院10个研究机构（表1）	学科中心
3	植物科学数据中心	中国科学院植物研究所	学科中心
4	地球大数据科学数据中心	中国科学院空天信息创新研究院	学科中心
5	海洋科学数据中心	中国科学院海洋研究所	学科中心
6	凝聚态物质科学数据中心	中国科学院物理研究所	学科中心
7	脑科学数据中心	中国科学院脑科学与智能技术卓越创新中心	学科中心
8	化学化工科学数据中心	中国科学院过程工程研究所	学科中心
9	干细胞与再生医学科学数据中心	中国科学院动物研究所	学科中心

对孤立。过去20年间，国家重点鼓励创建数据门户，重点解决科学数据的“可检索、可浏览、可共享”的共性基础问题，在开放数据公共平台和配套设施建设方面取得了较大的进展。但总体来看，在促进数据重用、促进科研创新和社会开放创新等方面仍面临巨大挑战[43]。具体存在4个方面的问题。

（1）当前建设规模不能满足日益增长的数据管理和使用的需要。从建设规模和财政资金投入而言，目前较为重视的仍然是大科学装置层级或国家级数据基础设施的建设，还无法满足整个科技界乃至社会对开放科学数据管理和使用的需要。

（2）现有标准体系和技术能力仍无法满足建设需求[36]。目前，针对大数据管理和处理标准、算法及工具层出不穷，但建设满足各类需求的开放数据基础设施挑战仍然很大，包括：① 缺乏有效的标准体系和查询手段；② 缺乏对体系架构的标准化建模，导致不同学科不同行业的数据很难重用，系统间集成和互通困难；③ 缺乏对数据管理标准化环节的重视；④ 缺乏有效的标准使用和升级指南，导致历史遗留系统和新系统之间的兼容较为困难。

（3）建设模式单一，支持数据整合、数据分析和支撑科技决策的功能较弱。当前，国内开放数据基础设施的资金来源单一，跨部门跨领域的合作少，存在自成体系或者重复建设的情况。开放数据基础设施目前主要支持对同领域同类型的数据整合功能，而对不同领域来源的不同类型的数据集进行整合、协同数据分析和支撑科技决策的功能较弱。

（4）在科学数据长期保存和再利用方面缺乏整体规划。我国大科学装置产生的科学数据规模巨大，长期利用价值高。尤其是在大科学装置或大科学项目结束运行后，需要有效的数据保存与管理、软件与计算支撑、完善的技术档案来确保科学数据的长期可用。目前除少数领域开始研究制定长期保存和再利用的规划外，我国还缺乏相关的整体规划。

除上述问题外，我国对开放数据基础设施建设的政策尚缺乏系统性设计和全方位考虑，各级政府、各类机构发布的相关政策意见主要侧重于“硬”条件的建设，而对“软”内容建设的重视和支持不足。

3 加强我国开放数据基础设施建设的思考与建议

开放科学数据需要强大的、可持续的基础设施和健全的政策制度支持，开放数据基础设施的实施路径应以充分挖掘数据价值为目标，使数据“存得下、流得动、用得好”。针对我国开放数据基础设施建设，提出以下4点建议。

3.1 加强顶层设计，统一规划布局开放数据基础设施建设，建立综合性的国家数据中心和国际数据中心

开放数据基础设施外延广、内涵丰富，需要加强国家层面的顶层设计和实施路径规划，以保障开放数据基础设施政策制定的连贯性和可操作性。

建议：① 开放数据基础设施的建设应以开放数据平台建设为核心。开放数据平台以数据为主体，通过高度融合存储、计算、网络和软件资源，实现开放数据的最大价值的挖掘。一方面，统筹开放数据平台建设，研究确立开放数据平台的整体框架、服务体系、认证标准和评价机制；另一方面，重视开放数据平台的可持续发展，从制定差异化数据政策、提供数据采集和使用的评价、提供支持服务3个方面入手，进一步优化投入机制，鼓励引导不同创新主体参与数据发展，形成以国家、部委和国家数据中心运行机构投入为主导，多元化投入相结合的数据资源建设和服务运行的资金保障体系，确保开放数据平台的可持续发展。② 开放数据基础设施建设有助于打破数据壁垒。我国现有开放数据基础设施主要集中在各类学科数据中心或共享平台，不能适应以大数据、物联网、人工智能为代表的新技术革命蓬勃发展，无法加速多领域

科学数据交叉应用和向现实生产力的转化。建设综合性基础性国家数据中心和国际数据中心是必然的解决途径。可持续发展大数据国际研究中心已经进行了先行探索，并取得了显著成效，为推动我国的开放数据基础设施的跨越式发展提供了经验积累。

3.2 坚持合作开放的科学数据基础设施云建设-云安全-云应用的实施路径

开放数据基础设施能有效提升科研的效率、参与度与可见性，加强科研质量与严谨度，促进科研团队跨学科合作。2019年，在CODATA北京会议上，中国科学家代表提出协作共建“全球开放科学云”（GOSC）的倡议，目前已与全球主要信息基础设施和国际组织、平台达成广泛共识并建立定期对话机制，研发构建了首个中欧跨洲际云联邦实验床。

建议：未来，我国应在国际大科学计划、大科学装置的牵引和推动下，充分发挥我国具有明显优势的数据资源的主导作用，着力推动相关数据分析方法及工具集合的研发，加强与国际组织、国家之间的沟通交流和培训，积极参与并推动国际层面的数据共享与合作应用；同时，应充分发挥中国科学院开放基础设施的引领和推动作用，合作建立国际共享的开放数据云服务体系。

3.3 营造融通数据生态，发展基于开放科学的创新驱动发展的范式

以开放数据获取为前导发展起来的开放科学和公民科学，与可持续科学密切相关。面向发展与合作，科技创新应该主要体现在将科技进步用于创造新需求、新应用、新业态和新市场，同时创新科技合作模式。

近年，欧美等国新布局的一批科学数据基础设施呈现问题导向、打破科学数据学科领域边界的趋势。例如，融合神经科学、分布式计算技术的欧洲脑研究基础设施；支持能源消耗和绿色交易的欧洲计算/通信实验大型研究基础设施等。这些设施有望充分调动多领域数据资源，融合自然科学各领域、自然科学与社会科学，成为推进交叉融合和跨领域互操作的先行者，营造跨学科、跨尺度、跨时空的科学数据共享生态。

建议：鼓励发起和建设科学、社会问题导向的科学数据基础设施，调动盘活多学科、跨领域的科学数据资源，为解决大规模、复杂性科学问题和社会挑战提供支撑。

3.4 发展基于开放科学的创新驱动发展的范式，重塑国际科技合作治理模式

开放科学有利于新科技革命和产业变革突破经典技术极限，形成新规则、新政策、新评估标准和新指标体系。开放科学通过对科学技术的兼收并蓄，耦合了不同地域、不同领域、不同团队的科研优势，利于促进全人类都能够站在巨人肩膀上做科研，形成累积效应。开放科学数据是实现开放科学的基本条件之一。秉承开放科学理念，有利于迅速提升我国科研实力。

开放数据基础设施作为开放科学数据的载体，可为政策制定者提供更为全面的综合数据和信息，为全球性挑战的综合政策制定提供方案，同时也可为科研人员的跨领域、跨区域合作提供新路径。

建议：遵循从开放数据到开放科学的发展潮流，利用云计算、大数据、区块链等先进技术与方法，将人工智能赋能大数据，打造集数据-计算-服务于一体的数据共享新模式，促进多学科数据关联分析和信息融合，深化多领域数据的综合应用，驱动重大科学发现与决策支持。

参考文献

1 盛小平，毕畅畅，唐筠杰. 国内外开放科学主题研究综述. 图书情报知识, 2022, 39(4): 101-113.
Sheng X P, Bi C C, Tang J J. A review of open science research at home and abroad. Documentation, Information & Knowledge, 2022, 39(4): 101-113. (in Chinese)

2 Ramachandran R, Bugbee K, Murphy K. From open data to open science. Earth and Space Science, 2021, 8(5): e2020EA001562.

3 潘小多, 李新, 冉有华, 等. 开放科学背景下的科学数据开放共享：国家青藏高原科学数据中心的实践. 大数据, 2022, 8(1): 113-120.
Pan X D, Li X, Ran Y H, et al. Open access of scientific data in the context of open science: The practice of the National Tibetan Plateau Data Center. Big Data Research, 2022, 8(1): 113-120. (in Chinese)

4 范英杰, 樊春良. 寻求共同基础 推进交流合作——对美国智库和科学界主要科技政策报告的解读与启示建议. 中国科学院院刊, 2022, 37(2): 197-205.
Fan Y J, Fan C L. Seeking common ground to promote exchanges and cooperation—Interpretation, enlightenment and suggestions of major science and technology policy reports of American think tanks and scientific communities. Bulletin of Chinese Academy of Sciences, 2022, 37(2): 197-205. (in Chinese)

5 United Nations Educational, Scientific, and Cultural Organization. UNESCO recommendation on open science. (2021-11-23)[2022-12-20]. https://unesdoc.unesco.org/ark:/48223/pf0000379949.

6 王菲, 平安. 开放获取环境下国家图书馆外文信息资源建设的思考. 国家图书馆学刊, 2022, 31(4): 93-101.
Wang F, Ping A. Analysis of the foreign information resource development in the national library of China in the context of open access. Journal of the National Library of China, 2022, 31(4): 93-101. (in Chinese)

7 郭华东, 王力哲, 陈方, 等. 科学大数据与数字地球. 科学通报, 2014, 59(12): 1047-1054.
Guo H D, Wang L Z, Chen F, et al. Scientific big data and digital Earth. Chinese Science Bulletin, 2014, 59(12): 1047-1054. (in Chinese)

8 Borowitz M. The Global Effort for Open Access to Environmental Satellite Data. Cambridge: The MIT Press, 2017.

9 盛小平, 杨智勇. 开放科学、开放共享、开放数据三者关系解析. 图书情报工作, 2019, 63(17): 15-22.
Sheng X P, Yang Z Y. Analysis of the relationship among open science, open sharing and open data. Library and Information Service, 2019, 63(17): 15-22. (in Chinese)

10 科学数据共享调研组. 科学数据共享工程的总体框架. 中国基础科学, 2003, 5(1): 63-68.
Scientific Data Sharing Research Group. Overall framework of scientific data sharing project. China Basic Science, 2003, 5(1): 63-68. (in Chinese)

11 陈晓玲, 全志薇, 毛刚, 等. 开放共享视角下科学数据研究综述. 图书情报导刊, 2022, 7(5): 72-79.
Chen X L, Quan Z W, Mao G, et al. Review of scientific data research from the perspective of open sharing. Journal of Library and Information Science, 2022, 7(5): 72-79. (in Chinese)

12 赵瑞雪, 赵华, 郑建华, 等. 科研机构科学数据管理实践与展望. 农业大数据学报, 2019, 1(4): 65-75.
Zhao R X, Zhao H, Zheng J H, et al. Scientific data management in scientific research institutions: Practice and prospects. Journal of Agricultural Big Data, 2019, 1(4): 65-75. (in Chinese)

13 阿儒涵, 吴丛, 李晓轩. 科研数据开放的国际实践及对我国的启示. 中国科学院院刊, 2020, 35(1): 11-18.
Aruhan, Wu C, Li X X. International practice of open research data and its enlightenment to China. Bulletin of Chinese Academy of Sciences, 2020, 35(1): 11-18. (in Chinese)

14 Committee on Data for Science and Technology. The Beijing Declaration on Research Data. Paris: CODATA, 2019.

15 Wilkinson M D, Dumontier M, Aalbersberg I J, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 2016, 3: 160018.

16 邢文明, 郭安琪, 秦顺, 等. 科学数据管理与共享的FAIR原则——背景、内容与实施. 信息资源管理学报, 2021, 11(2): 60-68.
Xing W M, Guo A Q, Qin S, et al. FAIR principle of scientific data management and sharing: Background, content and implementation. Journal of Information Resources Management, 2021, 11(2): 60-68. (in Chinese)

17 Carroll S R, Herczog E, Hudson M, et al. Operationalizing the CARE and FAIR Principles for Indigenous data futures.

Scientific Data, 2021, 8: 108.

18 孙九林, 黄鼎成, 李晓波. 我国科技数据管理和共享服务的新进展. 世界科技研究与发展, 2002, 24(5): 15-19.
Sun J L, Huang D C, Li X B. The new progress of science data managing and sharing service in China. World Sci-Tech R & D, 2002, 24(5): 15-19. (in Chinese)

19 张丽丽, 温亮明, 石蕾, 等. 国内外科学数据管理与开放共享的最新进展. 中国科学院院刊, 2018, 33(8): 774-782.
Zhang L L, Wen L M, Shi L, et al. Progress in scientific data management and sharing. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 774-782. (in Chinese)

20 国务院办公厅. 国务院办公厅关于印发科学数据管理办法的通知. (2018-03-17)[2023-05-25]. https://most.gov.cn/xxgk/xinxifenlei/fdzdgknr/fgzc/gfxwj/gfxwj2018/201804/t20180404_139023.html.
General Office of the State Council of the People's Republic of China. Scientific Data Management Measures. (2018-03-17)[2023-05-25]. https://most.gov.cn/xxgk/xinxifenlei/fdzdgknr/fgzc/gfxwj/gfxwj2018/201804/t20180404_139023.html. (in Chinese)

21 郭华东. 大数据 大科学 大发现——大数据与科学发现国际研讨会综述. 中国科学院院刊, 2014, 29(4): 500-506.
Guo H D. Big data, big science, big discovery—Review of CODATA workshop on big data for international scientific programmes. Bulletin of Chinese Academy of Sciences, 2014, 29(4): 500-506. (in Chinese)

22 王卷乐, 孙九林. 世界数据中心（WDC）回顾、变革与展望. 地球科学进展, 2009, 24(6): 612-620.
Wang J L, Sun J L. Review, Reform and prospect analysis of world data center. Advances in Earth Science, 2009, 24(6): 612-620. (in Chinese)

23 郭华东. 科学大数据——国家大数据战略的基石. 中国科学院院刊, 2018, 33(8): 768-773.
Guo H D. Scientific big data—a footstone of national strategy for big data. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 768-773. (in Chinese)

24 郭华东. 地球大数据科学工程. 中国科学院院刊, 2018, 33(8): 818-824.
Guo H D. A project on big earth data science engineering. Bulletin of Chinese Academy of Sciences, 2018, 33(8): 818-824. (in Chinese)

25 Guo H, Goodchild M F, Annoni A. Manual of Digital Earth. Singapore: Springer, 2020.

26 郭华东, 梁栋, 陈方, 等. 地球大数据促进联合国可持续发展目标实现. 中国科学院院刊, 2021, 36(8): 874-884.
Guo H D, Liang D, Chen F, et al. Big earth data facilitates sustainable development goals. Bulletin of Chinese Academy of Sciences, 2021, 36(8): 874-884. (in Chinese)

27 Guo H D, Liang D, Chen F, et al. Innovative approaches to the Sustainable Development Goals using Big Earth Data. Big Earth Data, 2021, 5(3): 263-276.

28 侯建国. "地球大数据驱动联合国可持续发展目标实现"专刊序言. 中国科学院院刊, 2021, 36(8): 873.
Hou J G. Preface of special issue on big earth data for implementing the sustainable development goals. Bulletin of Chinese Academy of Sciences, 2021, 36(8): 873. (in Chinese)

29 刘凤红, 张恬. 开放科学背景下新兴学术论文出版类型——研究要素出版. 中国科技期刊研究, 2017, 28(2): 138-144.
Liu F H, Zhang T. An emerging type of scholarly articles in the Open Science background: Research elements publishing. Chinese Journal of Scientific and Technical Periodicals, 2017, 28(2): 138-144. (in Chinese)

30 关琳琳, 马瀚青, 王长林. 创办国际数据期刊的实践与思考——以*Big Earth Data*为例. 中国科技期刊研究, 2020, 31(1): 56-62.
Guan L L, Ma H Q, Wang C L. Practice and thoughts on establishing international data journals: A case study of *Big Earth Data*. Chinese Journal of Scientific and Technical Periodicals, 2020, 31(1): 56-62. (in Chinese)

31 Guo H D. Big data drives the development of Earth science. Big Earth Data, 2017, 1(1-2): 1-3.

32 王卷乐, 王玉洁, 张敏, 等. 2020年地球数据科学与共享热点回眸. 科技导报, 2021, 39(1): 105-114.
Wang J L, Wang Y J, Zhang M, et al. Review of the progress of earth data science and sharing in 2020. Science & Technology Review, 2021, 39(1): 105-114. (in Chinese)

33 郭华东, 陈润生, 徐志伟, 等. 自然科学与人文科学大数

据——第六届中德前沿探索圆桌会议综述. 中国科学院院刊, 2016, 31(6): 707-716.

Gou H D, Chen R S, Xu Z W, et al. Big data in natural sciences, humanities and social sciences—Review of the 6th exploratory round table conference. Bulletin of Chinese Academy of Sciences, 2016, 31(6): 707-716. (in Chinese)

34 滕吉文, 刘有山, 皮娇龙. 科学与技术的发展与大数据的时代反响——地球科学新信息的获取与创新再造. 地球物理学进展, 2016, 31(1): 1-22.

Teng J W, Liu Y S, Pi J L. Development of science and technology, times echo of big data—Acquisition, innovation and reconstruction of new information of earth sciences. Progress in Geophysics, 2016, 31(1): 1-22. (in Chinese)

35 汪洋, 周园春, 王彦棡, 等. 适度超前推动科研基础平台建设支撑我国高水平科技自立自强. 中国科学院院刊, 2022, 37(5): 652-660.

Wang Y, Zhou Y C, Wang Y G, et al. Promoting infrastructure construction in advance to support sci-tech self-reliance and self-strengthening at higher level. Bulletin of Chinese Academy of Sciences, 2022, 37(5): 652-660. (in Chinese)

36 赵志明, 廖晓锋, 王晓玲, 等. 农业大数据基础设施开发的参考模型方法. 华东师范大学学报（自然科学版）, 2019, (2): 77-96.

Zhao Z M, Liao X F, Wang X L, et al. A Reference Model approach for developing agricultural big data infrastructures. Journal of East China Normal University (Natural Science), 2019, (2): 77-96. (in Chinese)

37 西桂权, 付宏, 刘光宇. 中国大科学装置发展现状及国外经验借鉴. 科技导报, 2020, 38(11): 6-15.

Xi G Q, Fu H, Liu G Y. Current development of large scale scientific facilities in China and foreign experiences. Science & Technology Review, 2020, 38(11): 6-15. (in Chinese)

38 王贻芳. 我国重大科技基础设施的现状和未来发展. (2022-06-24)[2022-12-02]. http://www.npc.gov.cn/npc/c30834/202206/90acffdbc88045758409c17e7dd86971.shtml.

Wang Y F. The Current Situation and Future Development of Major Science and Technology Infrastructure in China. (2022-06-24)[2022-12-02]. http://www.npc.gov.cn/npc/c30834/202206/90acffdbc88045758409c17e7dd86971.shtml. (in Chinese)

39 罗小安, 杨春霞. 中国科学院重大科技基础设施建设的回顾与思考. 中国科学院院刊, 2012, 27(6): 710-716.

Luo X A, Yang C X. Retrospect and consideration on construction of large research infrastructures of CAS. Bulletin of Chinese Academy of Sciences, 2012, 27(6): 710-716. (in Chinese)

40 Guo H D, Liang D, Sun Z C, et al. Measuring and evaluating SDG indicators with Big Earth Data. Science Bulletin, 2022, 67(17): 1792-1801.

41 邱春艳. 国内外科学数据相关实践中的元数据研究进展. 情报资料工作, 2021, 42(5): 104-112.

Qiu C Y. Metadata research progress in scientific data-related practices at home and abroad. Information and Documentation Services, 2021, 42(5): 104-112. (in Chinese)

42 李洋, 温亮明. 《科学数据管理办法》落实现状、影响因素及推进策略研究. 图书情报工作, 2021, 65(2): 65-74.

Li Y, Wen L M. Research on the implementation status, influencing factors and promotion strategies of the “Scientific Data Management Rule”. Library and Information Service, 2021, 65(2): 65-74. (in Chinese)

43 Li X, Cheng G D, Wang L X, et al. Boosting geoscience data sharing in China. Nature Geoscience, 2021, 14(8): 541-542.

Strengthening Open Data Infrastructure and Promoting Open Science

GUO Huadong[1,2*] CHEN Hesheng[3] YAN Dongmei[1,2] LI Jianhui[1,4] CHEN Gang[3]
HE Guojin[1,2] MA Juncai[5] LIANG Dong[1,2] KONG Lingyi[1,2]
（1 International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China;
2 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;
3 Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China;
4 Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China;
5 Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China）

Abstract To promote economic development, social progress, and scientific and technological innovation, it is necessary to strengthen scientific cooperation and information sharing. Open data has emerged in response and become a seemingly inevitable development in the evolution of digital technology. Open data, however, must be supported by infrastructure composed of physical entities and virtual systems that meet the needs of data applications in many fields. Constructing and strengthening open data infrastructure should therefore be considered important objectives of information technology development. This study analyzes the elements of open data infrastructure and expounds its significant positive role in implementing open science. Based on an analysis of the current state and substantial development of China's open data infrastructure, this study puts forward relevant measures and suggestions in view of the shortcomings and challenges China has faced with its open data infrastructure.

Keywords open data, open data infrastructure, open science, Big Earth Data, scientific data center

郭华东 中国科学院院士、俄罗斯科学院外籍院士、芬兰科学与人文院外籍院士、发展中国家科学院院士、国际科学理事会会士。可持续发展大数据国际研究中心主任。主要从事遥感信息科学、雷达对地观测、数字地球等领域研究。现任国际数字地球学会名誉主席、联合国教科文组织国际自然与文化遗产空间技术中心主任、《国际数字地球学报》和《地球大数据》主编等。E-mail: hdguo@radi.ac.cn

GUO Huadong Director General of the International Research Center of Big Data for Sustainable Development Goals. He is an Academician of the Chinse Academy of Sciences, a Foreign Member of the Russian Academy of Sciences, a Foreign Member of the Finnish Society of Sciences and Letters, a Fellow of TWAS, and a Fellow of ISC. He specializes in remote sensing, radar for Earth observation, and Digital Earth science. He presently serves as Honorary President of the International Society for Digital Earth, Director of the International Center on Space Technologies for Natural and Cultural Heritage under the Auspices of UNESCO, and Editor-in-Chief of two scientific journals *International Journal of Digital Earth* and *Big Earth Data*. E-mail: hdguo@radi.ac.cn

■责任编辑：张帆

*Corresponding author